Modified Nearest Neighbor Untuk Prediksi Curah Hujan

Muhammad[1)] , Bambang Lareno[2)]
[1]Teknik Informatika, STMIK PPKIA Tarakanita Rahmawati Tarakan
[2]Teknik Informatika, STMIK Indonesia Banjarmasin
e-mail: [1]besjem@yahoo.com, [2]blareno@gmail.com

Abstrak

Telah banyak penelitian mengenai prediksi curah hujan berbasis neural network, dan umumnya memakai data tahunan. Belum diketahui bagaimana jika menggunakan algoritma klasifikasi yang dimodifikasi untuk menganalisa data tersebut, seperti Nearest Neighbor. Berdasarkan hal tersebut, maka penelitian ini bertujuan untuk menguji akurasi prediksi curah hujan algoritma berbasis Nearest Neighbor dan membandingkannya dengan hasil prediksi Back Propagation Neural Network. Dari hasil penelitian dapat disimpulkan bahwa akurasi terbaik untuk prediksi 12 bulan, dihasilkan oleh BPNN–lm, 82,46%. sedangkan untuk prediksi 24 bulan, MAPE terbaik dihasilkan BPNN-lm. Sedangkan RMSE dan MAD, dihasilkan oleh BPNN-scg. Algoritma KNN belum dapat memprediksi lebih akurat dari pada BPNN Levenberg-Marquardt dalam memprediksi curah hujan. Penerapan KNN memiliki nilai lebih, yaitu memperlebar area penerapan algoritma klasifikasi NN yaitu pada prediksi timeseries curah hujan.

Kata kunci: *timeseries, sequence, suffix, manhattan-distance*

1. Pendahuluan

Prakiraan curah hujan, sebagai salah satu data klimatologi, mempunyai peran yang penting sebagai salah satu bahan pertimbangan bagi pembuat keputusan, khususnya dalam lingkungan kementerian terkait. Hal ini karena informasi besaran curah hujan mempunyai nilai ekonomi dalam berbagai kegiatan, mulai dari pertanian sampai dengan pengendalian banjir. Selain itu pencatatan curah hujan yang akurat, rapi dan periodik dapat digunakan untuk analisa pola iklim[1]. Jadi data curah hujan termasuk data rentet waktu (*time-series*), sehingga dapat dianalisa dan diprediksi dengan pendekatan neural network[2].

BMKG telah mengupayakan pendekatan dengan statistik, *wavelet*, ANFIS dan Tisean dalam bentuk aplikasi HyBMKG untuk prakiraan curah hujan[3]. Gumarang, dkk menggunakan BPNN untuk mengestimasi curah hujan di kota Pontianak[4]. Charaniya tertarik untuk merancang model neural network untuk prediksi hujan daerah Nagpur, India. Namun pola dan intensitas hujan India cukup berbeda dengan Indonesia[5]. Budi Warsito telah menerapkan algoritma *neural network* dengan *learning* Levenberg-Marquardt untuk memprediksi curah hujan kota Semarang[6]. Data yang dipergunakan adalah data curah hujan bulanan kota Semarang.

Dari latar ini dapat diambil kesimpulan bahwa telah banyak penelitian mengenai prediksi curah hujan berbasis neural network, dan biasanya memakai data tahunan. Sedangkan jika memakai data bulanan, sering hanya dalam kurun 10 tahun atau dengan kata lain hanya 120 baris data. Belum diketahui bagaimana jika menggunakan algoritma klasifikasi yang dimodifikasi untuk menganalisa data rentet waktu, seperti Nearest Neighbor. Selain itu, bagaimana hasil akurasi jika menggunakan lebih dari 120 baris data.

Berdasarkan latar belakang di atas, maka perlu untuk menguji akurasi prediksi k-Nearest Neighbor curah hujan. Sehinggga Question Research: "Bagaimana mengevaluasi akurasi algoritma Nearest-Neighbor curah hujan dan membandingkannya dengan hasil algoritma neural network, sehingga dapat diketahui mana yang lebih akurat?"

2. Metode Penelitian

Penelitian ini memakai data metode penelitian eksperimen komparasi, yang terdiri: (1)Metode Pengumpulan data dan pengolahan data awal, (2)Metode yang diusulkan, (3)Eksperimen dan pengujian model, (4)Hasil eksperimen dan (5)Evaluasi dan validasi hasil.

2.1. Pengumpulan Data

Penelitian ini memakai data curah hujan dan kelembaban yang didapatkan dari BMKG Kalimanatan Barat. Data yang dibutuhkan dalam penelitian ini adalah: data curah hujan (mm) bulanan periode 1990 – 2012 (23 tahun) dari Stasiun Klimatologi Supadio BMKG Provinsi Kalimantan Barat.

2.2. Pengolahan Data Awal

Data yang didapatkan dari instasi terkait masih berupa data curah hujan harian, sehingga harus direkapitulasi terlebih dahulu. Rekapitulasi tersebut dilakukan dengan memperhatikan kebutuhan. Data yang digunakan (untuk BPNN) kemudian ditransformasi sehingga berada dalam rentang 0-1, transformasi ini tidak mengubah pola data, hanya besarannya saja. Sedangkan data yang digunakan untuk KNN tidak ditransformasi, cukup di sesuaikan dengan panjang sequence data dan jumlah k yang diinginkan.

2.3. Metode/Model yang diusulkan

Metode yang digunakan adalah perbandingan antara akurasi yang dihasilkan oleh pelatihan dengan data curah hujan bulanan. Pelatihan data menggunakan Algoritma K-Nearest Neighbor (KNN) dan Algoritma Backpropagation Neural Network (BPNN).

2.3.1. Backpropagation Neural Network (BPNN)

BPNN diaplikasikan menggunakan Matlab 2009b dengan algoritma learning yang berbeda: Algoritma Learning Levenberg Marquardt (trainlm) dan Algoritma Learning Scaled Conjugate Gradient (trainscg). Algoritma BPNN akan diterapkan pada data curah hujan bulanan melalui suatu model simulasi. Evaluasi dilakukan dengan mengamati hasil prediksi dibandingkan dengan data sebenarnya.

2.3.2. Nearest Neighbor

Untuk yang langkah pertama, data {x1, x2,. . . , X_n} menjadi rentetan yang akan dianalisis. Setelah itu, dicari panjang subsequences k (anggap sepanjang ws) dalam rentet data yang dianalisis, yang paling dekat dengan akhiran vektor subsequence (suffix) {x_n-ws + 1,. . . , X_n}, dengan ws> 1, adalah tetangga terdekat terbaik[7]. k-subsequences terbentuk dalam:

$$\{x_{q1}, x_{q1+1}, \dots, x_{q1+ws-1}\}, \dots, \{x_{qk}, x_{qk+1}, \dots, x_{qk+ws-1}\} \quad (1)$$

di mana $1 \leq qi \leq$ n-ws, i = {1, 2,. . . , k}. Ukuran Lp dapat digunakan untuk menghitung jarak antara akhiran tiap utaian dan tetangganya. Nilai-nilai k dan ws adalah parameter dari model prediksi[8].

Ada dua pendekatan untuk langkah kedua di atas. Pendekatan pertama yaitu pekerjaan Giles, dkk[9]yang fokus memprediksi arah variasi untuk nilai berikutnya. Ada tiga hal yang mereka telaah: (1)meningkatkan tren, (2)tren menurun dan (3) nilai konstan. Mereka menganggap bahwa kecenderungan akan konstan ketika variasi yang ada tidak melebihi nilai (persen) yang telah ditetapkan. Pendekatan kedua didasarkan pada menghitung nilai prediksi[10][11]. Penelitian ini mengikuti pendekatan kedua.

Jadi, untuk memprediksi nilai p berikutnya, perlu menghasilkan nilai {x_{n+1},. . . , X_{n+p}}. Proses ini merupakan langkah berulang, setiap langkah akan memprediksi nilai xn berikutnya x_{n+i} ($1 \leq i \leq p$).

Untuk memprediksi nilai x_{n+1}, perlu perhitungan posisi relatif sequence tetangga terdekat terpilih, sebut saja $\{x_a, x_{a+1}, \dots, x_{a+ws-1}\}$ dengan sequence dari suffix $\{x_{n+i-ws}, \dots, x_{n+i-1}\}$. Jarak (Δ) didapatkan dari selisih data pertama sequence terpilih dengan data pertama suffix:

$$\Delta_a = x_{n+i-ws} - x_a \quad (2)$$

Sehingga dengan mengadopsi jarak *Manhattan* didapat jarak relatif (d):

$$d = \sum_{j=0}^{ws-1} |x_{n+i-ws+j} - x_{a+j} - \Delta_a| \quad (3)$$

Dengan demikian, didapatkan:

$$x_{n+i} = x_{q_t^i + ws} + \Delta_a \quad (4)$$

Persamaan (2) → (4):

$$x_{n+i} = x_{q_t^i + ws} + x_{n+i-ws} - x_{q_t^i} \quad (5)$$

Algoritma KNN yang digunakan dalam bentuk pseudocode sebagai berikut[12]:

Input data curah hujan secara series { $x_1, x_2, \dots, x_n$}

For i = 1 to p

Cari sequence terdekat:

$\{ x_{q1}, x_{q1+1}, \ldots, x_{q1+ws-1}\}, \ldots, \{ x_{qk}, x_{qk+1}, \ldots, x_{qk+ws-1}\}$
dengan sequence suffix:
$\{ x_{n+i-ws}, \ldots, x_{n+i-1}\}$
Kemudian hitung dan perkirakan nilai x_{n+1} berdasarkan nilai
$x^{i}_{q1+ws}, \ldots, x^{i}_{qk+ws}$
Kemudian tambahkan nilai x_{n+1} sebagai data terakhir $\{ x_1, \ldots, x_{n+i-1}\}$

Next i

Output Series hasil prediksi $\{ x_{n+i}, \ldots, x_{n+p}\}$

Sebagai contoh, misalkan ada data yang berasal dari data hujan selama 60 bulan. Panjang ws yang ditentukan = 5, maka akan didapatkan k=9 dengan satu suffix.. Artinya sequence yang akan menjadi tetangga terdekat adalah subsequence 1 sampai 9 (perhatikan Tabel 1), yaitu { xq1, x q1+1, … , xq1+ 5-1}, … , { xq9, xq9+1, … , xq9+ 5-1} atau {117, 317, … , 249.9}, … , {326.1, … , 213} dengan suffix {68.1, …, 213}

Tabel 1. Data Sequence panjang 5 bulan (KNN-5) dengan k=9

k=9	1	2	3	4	5	ws
Subsequence1	117,0	317,0	170,0	273,7	249,9	106,8
Subsequence2	247,7	63,9	351,7	274,8	423,9	438,6
Subsequence3	364,7	89,9	202,4	175,5	397,3	219,0
Subsequence4	76,6	27,5	65,3	172,7	349,7	316,2
Subsequence5	376,7	193,2	285,7	253,8	215,6	163,8
Subsequence6	221,4	86,5	236,1	245,6	329,7	247,5
Subsequence7	448,6	148,9	298,9	214,4	575,8	180,7
Subsequence8	289,2	94,8	171,9	270,2	398,4	207,9
Subsequence9	326,1	209,8	460,3	376,0	213,0	368,5
Suffix	68,1	74,8	10,3	292,5	414,2	332,3

Dalam proses pencocokan, misalnya subsequence4 terpilih sebagai sequence tetangga terdekat (persamaan 2 dan 3) maka nilai perkiraan x_{n+1} didapatkan (persamaan 6) dengan menambahkan nilai ws subsequence4 dengan nilai awal sequence, kemudian dikurang dengan nilai awal sub sequence4, x_{n+1} = 316,2 + 68,1 – 76,6 = 307,7.

Selisih absolut dengan data sebenarnya (error) = 332,3 – 307,7 = 24,6.

2.4. Eksperimen dan Pengujian Model/Metode

Kinerja neural network tergantung tidak hanya pada variabel input tetapi juga pada ukuran jaringan. Ukuran jaringan yang tidak memadai dan kesalahan pengaturan parameter akan mempengaruhi kecepatan konvergensi dan kualitas prediksi. Pemilihan nilai yang tepat dalam pengaturan parameter, sangat diperlukan karena banyak parameter harus dipilih dan ditetapkan sebelum proses pengujian.

Sebagian data digunakan sebagai data pelatihan untuk mendapatkan struktur terbaik BPNN, data 1990-2010 digunakan sebagai training dan checking, dengan komposisi 75:25. Data 2011-2012 digunakan sebagai data uji. Sedangkan untuk KNN, Data 1990-2010 digunakan untuk sequence tetangga terdekat dan data 2011-2012 digunakan sebagai data uji.

2.5. Evaluasi dan Validasi

Evaluasi dilakukan dengan mengamati hasil prediksi menggunakan algoritma BPNN dan KNN. Validasi dilakukan dengan mengukur hasil prediksi dibandingkan dengan data asal, sehingga akurasi masing-masing algoritma. Pengukuran kinerja dilakukan dengan menghitung error yang terjadi melalui besaran Percentage Error (MAPE), Root Mean Square Error (RMSE) dan Mean Absolute Deviation (MAD). Semakin kecil nilai error menyatakan semakin dekat nilai prediksi dengan nilai sebenarnya. Dengan demikian dapat diketahui algoritma mana yang lebih akurat dalam memprediksi curah hujan.

3. Hasil dan Pembahasan

3.1. Hasil Pengujian Model/Metode

3.1.1. BPNN

Dengan menggunakan Matlab proses pengujian dengan BPNN dilakukan. Hasilnya ditampilkan pada Tabel 2.

Tabel 2. Nilai RMSE Data Supadio

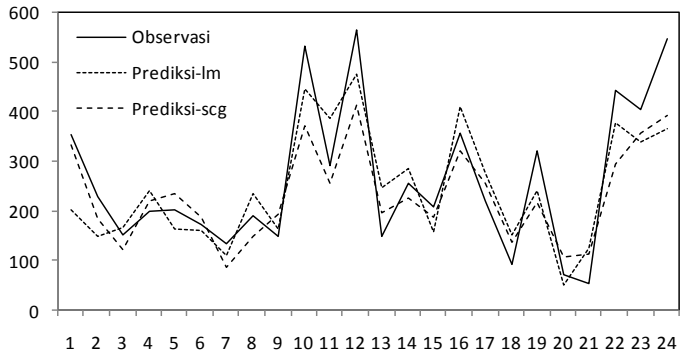
	Supadio		
	lm	scg	Rata-rata
	104,057	75,466	
	113,877	78,303	
	102,884	78,990	
	122,653	91,256	
	114,443	89,211	
	101,621	99,614	
	106,322	83,956	
	100,395	97,276	
	103,118	91,600	
	129,740	102,861	
MAX	129,740	102,861	116,300
MIN	100,395	75,466	87,931
AVG	109,911	88,853	99,382
MEDIAN	105,189	90,233	97,711
STDEV	9,965	9,476	9,721

Untuk mendapatkan error prediksi yang tervalidasi, data pisah mejadi 2 bagian. Data P1 sampai P263 untuk data pelatihan (data training), sedangkan data P241 sampai P264 untuk data uji. Berdasarkan RMSE hasil pelatihan pada arsitektur terbaik 4-4-1, pada Tabel 2, terlihat bahwa algoritma learning menggunakan scg lebih baik dari pada lm. Sedangkan untuk hasil data uji, yang ditampilkan pada tabel 3, menunjukkan bahwa learning lm lebih baik dari scg.

Tabel 3. Nilai Error Prediksi Data Uji 12 Bulan

Data Uji	Aktual	Prediksi-lm	Prediksi-scg	lm	scg	lm	scg	lm	scg
1	229,1	148,9	184,3	0,35	0,20	6432,0	2007,0	80,20	44,80
2	151,7	165,7	123,2	0,09	0,19	196,0	812,2	14,00	28,50
3	198,5	240,5	221,8	0,21	0,12	1764,0	542,9	42,00	23,30
4	204	165,2	235,2	0,19	0,15	1505,4	973,4	38,80	31,20
5	173,5	161,3	189,6	0,07	0,09	148,8	259,2	12,20	16,10
6	135,7	109,3	87,2	0,19	0,36	697,0	2352,3	26,40	48,50
7	190,1	233,9	148,3	0,23	0,22	1918,4	1747,2	43,80	41,80
8	147,9	164,4	194,4	0,11	0,31	272,3	2162,3	16,50	46,50
9	533,2	446,6	373,4	0,16	0,30	7499,6	25536,0	86,60	159,80
10	292,8	386,8	256,7	0,32	0,12	8836,0	1303,2	94,00	36,10
11	566,1	476,6	414,8	0,16	0,27	8010,3	22891,7	89,50	151,30
12	147,8	245,8	285,8	0,66	0,93	9604,0	19044,0	98,00	138,00
			MAPE	**17,44%**	**19,40%**				
			RMSE			**62,51**	**81,46**		
			MAD					**53,50**	**63,83**

Pada Tabel 3, terlihat bahwa perhitungan akurasi untuk data 12 bulan menghasilkan MAPE-lm 17,44% dan MAPE-scg 19,40%. RMSE-lm 62,51 dan RMSE-scg 81,46. MAD-lm 53,50 dan RMSE-scg 63,83. Jika dilanjutkan untuk data 24 bulan, perhitungan akurasi untuk data 24 bulan menghasilkan MAPE-lm 29,50% dan MAPE-scg 32,37%. RMSE-lm 79,19 dan RMSE-scg 76,93. MAD-lm 69,17 dan MAD-scg 63,23. Ditampilkan dalam bentuk grafik, terlihat sebagaimana gambar 1.



Gambar 1. Perbandingan hasil prediksi dengan data observasi Supadio

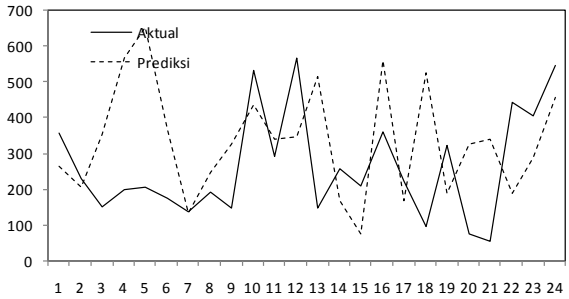
3.1.2 K Nearest Neighbor

Misalnya model data yang diperlukan KNN untuk data sequence 12 bulan dengan horison prediksi 1 bulan. Hasilnya ditampilkan pada Tabel 4

Tabel 4. Nilai Error Prediksi 12 Bulan KNN-12

Data Uji	Aktual	Prediksi	MAPE	RMSE	MAD
1	355,3	264,1	0,26	8317,4	91,20
2	229,1	206,8	0,10	497,3	22,30
3	151,7	352	1,32	40120,1	200,30
4	198,5	566,7	1,85	135571,2	368,20
5	204	653,4	2,20	201960,4	449,40
6	173,5	368,5	1,12	38025,0	195,00
7	135,7	134,4	0,01	1,7	1,30
8	190,1	246,9	0,30	3226,2	56,80
9	147,9	326	1,20	31719,6	178,10
10	533,2	435,7	0,18	9506,3	97,50
11	292,8	339,4	0,16	2171,6	46,60
12	566,1	345,8	0,39	48532,1	220,30
		Nilai Error Prediksi	**75,83%**	**208,10**	**160,58**

Dari tabel 4.6, terlihat bahwa perhitungan akurasi untuk data 12 bulan menghasilkan MAPE = 75,83%, RMSE = 208,10 dan MAD = 160,58. Sedangkan untuk horison prediksi 24 bulan, menghasilkan MAPE = 117,62%, RMSE = 219,94 dan MAD = 180,67. Secara grafik, ditunjukkan dengan gambar 2



Gambar 2. Perbandingan hasil prediksi mNN-12 untuk data uji 24 bulan

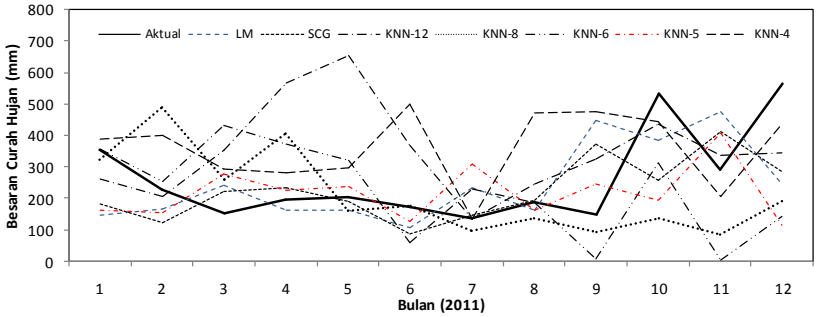
3.2. Evaluasi dan Validasi Hasil

Pada Tabel 5 terlihat bahwa error terkecil untuk prediksi 12 bulan, dihasilkan oleh BPNN–lm, yaitu MAPE 17,44% (akurasi 82,46%), Sedangkan untuk prediksi 24 bulan, MAPE terbaik dihasilkan BPNN-lm, 29,5% (atau akurasi 70,5%). Sedangkan RMSE dan MAD, dihasilkan oleh BPNN-scg.

Tabel 5. Nilai Agregat Akurasi

Parameter	Prediksi-lm	Prediksi-scg	KNN-12	KNN-8	KNN-6	KNN-5	KNN-4
12 Bulan							
MAPE	17,44%	19,40%	75,83%	52,18%	65,82%	29,53%	74,46%
RMSE	87,97	104,09	208,10	198,80	198,93	111,12	180,92
MAD	70,04	80,76	160,58	148,28	157,03	79,56	147,06
24 Bulan							
MAPE	29,50%	32,37%	117,62%	84,68%	59,26%	48,61%	78,08%
RMSE	79,19	76,93	219,94	149,68	122,58	148,58	128,15
MAD	69,17	63,23	180,67	159,73	140,46	111,92	148,24

Sedangkan untuk kNN, akurasi terbaik dihasilkan oleh kNN-5, yaitu sebesar 70,47% (MAPE 29,53%) untuk prediksi 12 bulan, sedangkan yang 24 bulan, sebesar 51,39% (MAPE 48,61%). Gambar 3 menunjukan pola hasil prediksi 12 bulan masing-masing algoritma.



Gambar 3. Perbandingan hasil prediksi untuk data uji 12 bulan

4. Simpulan

4.1. Kesimpulan

Dari hasil penelitian yang dilakukan dari tahap awal hingga pengujian, dan pengukuran, dapat disimpulkan bahwa Algoritma kNN belum dapat memprediksi lebih akurat dari pada BPNN Levenberg-Marquardt dalam memprediksi curah hujan. Akurasi terbaik yang hasilkan baru mecapai 70,47%. Akurasi terbaik untuk prediksi 12 bulan, dihasilkan oleh BPNN–lm, 82,46%. Sedangkan untuk prediksi 24 bulan, MAPE terbaik dihasilkan BPNN-lm. Sedangkan RMSE dan MAD, dihasilkan oleh BPNN-scg. Meski demikian, Penerapan kNN ini memiliki nilai lebih, yaitu memperlebar area penerapan algoritma klasifikasi NN yaitu pada prediksi timeseries curah hujan.

4.2. Saran

Berapa hal perlu disampaikan untuk pengujian yang lebih baik, yaitu: perlu pengujian dengan model folding, sehingga mengurangi efek data uji yang ekstrem. Perlu data tambahan dari wilayah lain yang memiliki pola hujan berbeda, khususnya yang memiliki pola hujan lokal. Selain itu, perlu data hujan yang lebih banyak sehingga masing-masing algoritma mendapatkan pelatihan yang cukup untuk mengenali pola hujan.

Daftar Pustaka

[1] Kadarsah SA. *Standardisasi Metadata Klimatologi Dalam Penelitian Perubahan Iklim Di Indonesia*. In Prosiding PPI Standardisasi 2010; 2010; Banjarmasin. p. 1-18.

[2] Gupta A, Gautam A, Jain C, Prasad H, Verma N. Time Series Analysis of Forecasting Indian Rainfall. *International Journal of Inventive Engineering and Sciences (IJIES)*. 2013 May; I(6): p. 42-45.

[3] Sonjaya I, Kurniawan T, Munir M, Wiratri M, Khairullah. Uji Aplikasi HyBMG Versi 2.0Untuk Prakiraan Curah HujanPola Monsunal Ekuatorial dan Lokal. *Buletin Meteorologi Klimatologi dan Geofisika*. 2009 September; 5(3): p. 323-339.

[4] Gumarang MI, Andromeda L, Nugroho BS. Estimasi Curah Hujan, Suhu dan Kelembaban Udara Dengan Menggunakan Jaringan Syaraf Tiruan Backpropagation. *Jurnal Aplikasi Fisika*. 2009 Februari; V(1): p. 54-61.

[5] Charaniya NA, Dudul SV. Design of Neural Network Models for Daily Rainfall Prediction. *International Journal of Computer Applications (0975 – 8887)*. 2013 January; 61(14): p. 23-27.

[6] Warsito B, Sumiyati S. Prediksi Curah Hujan Kota Semarang Dengan FeedForward NN Menggunakan Algoritma Quasi Newton BFGS dan Levenberg-Marquardt. 2005..

[7] Zhengzheng X, Jian P, Yu PS. Early Prediction on Time Series: A Nearest Neighbor Approach. *NSERC Discovery*. Simon Fraser University, 2010 .

[8] Guegan D, Rakotomarolahy P. The Multivariate k-Nearest Neighbor Model for Dependent Variables: One-Sided Estimation and Forecasting. Documents de travail du Centre d'Economie de la Sorbonne. Paris: Université Paris 1 Panthéon-Sorbonne, Paris School of Economics, CES-MSE; 2010. Report No.: ISSN : 1955-611X.

[9] Giles CL, Lawrence S, Tsoi AC. Noisy Time Series Prediction using Recurrent Neural Networks and Grammatical Inference. *Machine Learning*. Springer, 2001; 44(1-2): p. 161-183(23).

[10] Plummer EA. Time Series Forecasting with Feed-Forward Neural Networks: Guidelines and limitations; 2000.

[11] Yakowitz S. Nearest-neighbor methods for time-series analysis. *Journal of Time Series Analysis*. 1987; 8(2): p. 235-247.

[12] Sasu A. k-Nearest Neighbor Algorithm For Univariate Time Series Prediction. *Bulletin of the Transilvania University of Brasov*. 2012; 5(54) (2, Series III: Mathematics, Informatics, Physics): p. 147-152.